

Project 'Europeana Newspapers' gestart

Den Haag, 26 juni 2012



In het zojuist gestarte project Europeana Newspapers werkt een groep van 17 Europese instituten de komende drie jaar samen om meer dan 18 miljoen gedigitaliseerde krantenpagina's toe te voegen aan de online service Europeana. Europeana is het centrale toegangspunt tot miljoenen gedigitaliseerde boeken, schilderijen, films, museale objecten en archiefdocumenten uit heel Europa.

Het Europeana Newspapers project wordt gefinancierd door de Europese Commissie binnen het Kaderprogramma voor concurrentievermogen en innovatie (CIP) en is zowel gericht op aggregatie als op verrijking van het krantenmateriaal via The European Library.

Alle bibliotheken die deelnemen aan het project zullen hun full-text en gedigitaliseerde krantenmateriaal via Europeana verspreiden. Doel van het project is om het krantenmateriaal direct toegankelijk te maken voor gebruikers via een speciale interface in de content browser. Deze zal geïntegreerd worden in de Europeana portal en maakt het zoeken op zinsdelen of woorden in de krantentekst mogelijk. Dit is een grote stap vooruit vergeleken met de gebruikelijke zoekfunctionaliteit van bibliotheekcatalogi, waarin meestal alleen op datum of titel gezocht kan worden.

Het project zal daarnaast werken aan oplossingen voor diverse problemen op het gebied van gedigitaliseerde kranten, zoals Optical Character Recognition (OCR), Optical Layout Recognition (OLR), segmentatie van artikelen, Named Entity Recognition (NER) en pagina classificatie. OCR is de techniek waarmee gescande afbeeldingen van handgeschreven, getypte of geprinte tekst worden omgezet naar machine-leesbare tekst. Met OLR kunnen op een pagina met meerdere artikelen de afzonderlijke artikelen worden herkend en van elkaar gescheiden. NER maakt het mogelijk om specifieke namen in een tekst te herkennen en te classificeren volgens gestandaardiseerde namen van personen, plaatsen en organisaties.

Het project zal ook de kwaliteit van deze verrijkingstechnieken evalueren en de lokale metadata omzetten naar de Europeana Data Model standaard, in nauwe samenwerking met stakeholders uit de publieke en private sector.

Europeana Newspapers wordt gecoördineerd door de Staatsbibliothek zu Berlin – Preußischer Kulturbesitz. De voortgang van het project is te volgen via de website www.europeana-newspapers.eu. Voor meer informatie kunt u contact opnemen met Hans-Jörg Lieder of Thorsten Siegmann van de Staatsbibliothek zu Berlin via info@europeana-newspapers.eu.

Projectpartners:

Staatsbibliothek zu Berlin (coördinator)
National Library of Estonia
University of Helsinki, National Library of Finland
Bibliothèque nationale de France
CCS Content Conversion Specialists GmbH
National Library of Latvia
University of Beograd
Landesbibliothek Dr. Friedrich Tessimann
University of Salford

Koninklijke Bibliotheek
Österreichische Nationalbibliothek
Staats- und Universitätsbibliothek Hamburg
National Library of Poland
Stichting LIBER
National Library of Turkey
University of Innsbruck
The British Library
The European Library

Europeana in een notendop

Europeana is een meertalige online collectie van miljoenen gedigitaliseerde items uit Europese musea, bibliotheken, archieven en audiovisuele collecties. Momenteel biedt Europeana geïntegreerde toegang tot 23 miljoen boeken, films, schilderijen, museale objecten en archiefdocumenten van ruim 2.200 instellingen uit heel Europa.

www.europeana.eu

The Europeana Newspapers project is partially funded under the ICT Policy Support Programme (ICT PSP, http://ec.europa.eu/ict_psp) as part of the Competitiveness and Innovation Framework Programme by the European Community. This document reflects only authors' views. The EC is not liable for any use that may be done of the information contained therein.

